INFORMATION PROBLEMS FOR POLICY ANALYSIS AND FORECASTING

Robert S. Goldfarb
George Washington University

and

H.O. Stekler
George Washington University

"Unfortunately inferences drawn from sketchy data sometimes turn out to be wrong." *New York Times*, 18 April 1995.

Search theory contains a fundamental insight: information is not free for the price-searching consumer or for the job seeker. A similar finding also holds for the policy analyst or forecaster seeking "perfect information" about the structure of the economy. However, unlike the consumer or job seeker, the analyst may not be able to obtain "better" information with a marginal increase in effort or expenditure. In fact, the analyst may find that what *is* available is biased in subtle ways.

While there is general understanding of the information problems facing the consumer and job seeker, the analyst's problem of obtaining the requisite data has been less extensively investigated. Previous studies have identified a number of concrete difficulties in obtaining information. However, no study has comprehensively reviewed the *broad range* of information problems that might affect the analyst/forecaster.

Being aware of the informational pitfalls is important for both the analyst undertaking a policy study (or the forecaster generating predictions) and the user of that study. Poor information inputs may cause major errors in both analysis and interpretation. A prerequisite for minimizing these mistakes is to identify some of these potential sources of input error. This paper draws upon the literature to describe particular information problems that analysts might confront, and explores some of the implications. To focus the discussion, we address this hypothetical situation:

> A policy analyst or forecaster needs some basic information about the structure of a particular economic sector, or about how a proposed policy change might affect it. How might this basic information be obtained? What difficulties might be encountered?

This paper considers two distinctly different situations. In the first, information is needed to explain the current structure of an economic sector or to predict the behavioral responses of individuals or institutions under existing conditions. The concern here is with using *existing* data and knowledge for economic inference. In

Eastern Economic Journal, Vol. 22, No. 1, Winter 1996

the second, knowledge is required to predict the impact of policies that would change the economy's structure.

OBTAINING KNOWLEDGE OF THE CURRENT ECONOMIC STRUCTURE

Obtaining knowledge about the current structure and behavioral responses of the economy or its subsectors might seem relatively easy. Statistical agencies publish a mass of economic data. Moreover, innumerable published empirical studies use these data to describe aspects of the economy, or to test hypotheses. Finally, review articles integrate and summarize these empirical studies, and "meta-analyses" combine the findings of empirical studies to derive overall conclusions.

While the volume of published information is considerable, there are major problems in extracting the kind of basic initial knowledge needed to evaluate policies and make forecasts. The difficulties include problems with data revisions, the likelihood that published results conflict, and the possibility that the published literature may not reflect "true" knowledge because biases affect what is published.

Data Revision Problems

For statistical agencies to collect and process GDP data for any given quarter takes time. These agencies are thus confronted with the choice of delaying publication of GDP numbers until all the necessary data become available or releasing preliminary estimates to be revised later. Since it may take as long as three years to collect all of the data needed to estimate GDP, preliminary quarterly estimates are released 15 days after the end of that quarter and revised frequently.

These data revisions pose problems for forecasters. To make accurate predictions requires an understanding of the condition of the economy at the time that estimates of the future are generated. The GDP data published with very short lags may not, in fact, reflect actual conditions. Only future revisions reveal the "true" situation.[1] Moreover, estimation results may be sensitive to which data are used.[2] In short, the information available to analysts may be noisy and/or available only with a long lag.

Information from Published Studies

Problems may also affect the analyst's ability to draw information from published studies. Two important issues are (1) conflicting results and the methodology for deciding among the alternative views and (2) the possibility that what is published may be a biased sample of what is "true."

Koopmans [1979] suggested that any refutation-confirmation process involving results of quantitative economic studies is likely to be slow and diffuse. He was unable to find "in the literature a persuasive account of how such confirmation of premises can be perceived and documented...How do we keep the score of surviving hypotheses... (U)nresolved issues...drag on and remain unresolved" [1979, 12]. Koopmans's recognition that the process of "keeping track" or "keeping score" of empirical findings is itself problematical illustrates the difficulties of reconciling conflicting results.

Counting/Weighing Studies. The published empirical evidence available about a particular issue is often conflicting: studies may report quite different magnitudes or even directions of effects. In this case, the analyst is confronted with reconciling the results and determining which are "correct." It might be argued that "truth" can be obtained by "counting" or "weighing" how many studies come out on each side of the issue. If 70 percent of available studies favor one side of the issue, one might decide that the "true" information is contained in this subset of studies. Alternatively, one might attach a 70 percent probability that the results favor this side of the issue.

The procedure has some flaws. The analyst may be able to obtain only a small subset of the relevant studies and thus have difficulty keeping track of the results; some studies might be designed inappropriately; and some of the results might have been produced by older and less sophisticated techniques. These issues raise the question of determining appropriate weights for the various studies.

Meta-Analysis May Help. Meta-analysis has been described as "the use of formal statistical methods to combine the quantitative results of separate (but similar) studies" [Gelber and Goldhirsch, 1991, 461]; and as the "statistical analysis of... results from individual studies for the purpose of integrating the findings" [Cooper and Hedges, 1994, 537]. This technique has been applied extensively in the psychology/educational evaluation literature and in the biostatistics/medical research literature associated with clinical trials. Given its success in those fields, meta-analysis might also be useful in combining the results of conflicting empirical studies in economics.[3]

In conducting a meta-analysis, different statistical procedures are used to combine or synthesize results. When studies report the sizes of effects, the procedures are based on either ANOVA or regression approaches. When only probability levels of statistical significance are reported, different statistical tests can be used to combine results to obtain the overall significance level [Becker, 1994]. Finally, when only directional effects (signs of the outcomes) are reported, statistical tests may be applied to counts of directional or "sign" effects.

Meta-analysis can demonstrate that reliance on counts of directional or "sign" effects may sometimes produce misleading interpretations. Suppose 15 studies examined a particular relationship, with 11 showing a positive direction and the remaining four displaying a negative sign. With more than 70 percent of the results yielding the same positive effects, the temptation would be to believe that the effect was positive. However, application of the binomial test, assuming a 50-50 chance of a study showing either sign, would not be able to reject at the 5 percent level the null hypothesis that this 11-4 split occurred by chance. Thus, it would not be possible to infer from these studies (at this level of significance) that the effect was positive.

Meta-analysis provides techniques for the "weighing of the evidence" provided by a sizable empirical literature with possibly conflicting results. Thus, it might make a sizable contribution to determining how to make inferences from or "combine the results of" disparate studies. However, its application requires considerable care and sophistication.

Publication Biases. Publication bias represents another potential problem. Studies in a number of disciplines[4] suggest that even a careful reading of all the published literature might not reveal the true relationships. The problem is that the published literature may be a biased sample of all the relevant research which has been undertaken. Sources of bias stem from the possibility of "data mining" by the researcher, the researcher's submission strategies based on expectations about the type of manuscripts journals are likely to accept, and the chance that the behavior of editors might bias the selection of submitted studies actually published.

Even a relatively unanimous published literature could badly misrepresent the "true" population of confirmations versus refutations. This might be the case if a common mindset in doing empirical work leads the researcher to look for empirical results that support strongly held priors. For example, when empirical researchers have strong *a priori* expectations about the direction of effect ("sign") of particular variables, they might undertake additional statistical estimation until the results displayed the "correct" signs. That is, the "stopping rule" for running regressions is based on finding the "right" signs. But if some sizable proportion of published empirical work uses this kind of stopping rule, published studies will not be an unbiased sample of actual statistical results.[5] One possible example of this phenomenon is provided by Goldfarb [1995a]. He examined the literature investigating the effect that market concentration had on profits. He found that a change in theoretical expectations about what the data *should* show apparently resulted in actual changes in what researchers found empirically and what was actually published.[6]

The moral is that there is a serious danger that our initially-underinformed analyst might obtain misleading information from the published literature. *Caveat emptor.*

OBTAINING KNOWLEDGE FOR A CHANGING STRUCTURE

In this section we turn to the more difficult question of how to obtain and interpret information when the structure of the economy changes or a major new policy is to be, or has recently been, implemented. Typically, analysts try to determine the effect of structural changes by using theoretical and/or empirical models which have functioned well in the past to project into the future. This may be inadequate when fundamental structural changes are occurring, because such changes may render existing theories less applicable in ways that are hard to anticipate.

We first describe some problems in using empirical models. Second, we question whether theoretical analyses can "anticipate" the effects of structural change. As a case study, we consider airline deregulation, in which a theoretical framework was used to anticipate the structure and behavior of the deregulated industry. We provide an explanation for the discrepancies between the expected and observed outcomes in that industry.

Empirical Models

Two categories of empirical models have been used for projections into the future: large macroeconometric and computable general equilibrium models.

Macroeconometric Models. Many studies have described the major macroeconometric models and evaluated their short-run forecasts.[7] Someone interested in the effects of structural changes would be concerned primarily with the long-run dynamics of these models. The evidence is that different models yield disparate results when subjected to identical shocks. For example, across eleven models the range of the dynamic fiscal policy multiplier at the end of the third year, assuming fixed interest rates, was 1.29 to 7.93.[8]

But suppose that all existing forecasts of the dynamic fiscal multiplier (or other parameter of interest) could be reconciled, and a "best" point estimate obtained. Even using this consensus "best" estimate to anticipate the effects of possible structural change might be problematic. The Lucas critique [Lucas, 1976] postulated that these econometric models cannot provide valid projections for large structural changes or government interventions in the economy. The difficulty is that these changes might alter the behavioral responses embedded in the model. Consequently, the model's point forecasts could be questionable. The analyst must take these parameter uncertainties into account in interpreting simulation results.

Computable General Equilibrium Models. The models described above provide highly aggregated estimates, but the analyst might require more detail. In this case, a computable general equilibrium (CGE) model might be used to analyze the projected changes resulting from tax reform, NAFTA, and so forth. Is the information from such models likely to be valid?

In contrast to the large macro models, the coefficients of CGE models are typically not statistically estimated. Rather, the model parameters are obtained from other sources, including estimates from econometric studies, and adjusted to calibrate the model to a set of observed data assumed to reflect a general equilibrium.

Because the parameters are typically not statistically estimated, and the overall model has itself been calibrated to observed data, its fit is not readily subject to standard statistical tests.[9]

However, one can determine whether the parameters remain stable over time, and whether the results are robust with respect to alternative specifications. Many of the model's parameters are elasticities. Whalley [1985] points out that our assumptions about the sizes of these elasticities have changed drastically over the years and that the parameter estimates have *not* exhibited stability.[10]

In addition to examining the features of particular CGE models, it is also possible to compare the results from alternative CGE models. Coughlin [1990] showed that five different CGE models which examined the effects of the U.S.-Canada Free Trade Agreement yielded conflicting findings. Comparisons of CGE studies of the effects of NAFTA also showed that the results of different CGE analyses differed substantially.[11]

Given these results, how can the CGE models provide the analyst with useful information? One possibility is to restrict their use to situations that are more abstract, hypothetical, and generalized than a concrete comparison of (say) the detailed efficiency and distributional implications of actual trade regimes or tax reforms in all their complexity. This "backing-off" is consistent with Krueger's observation that good

CGE analyses make use of economic intuition, quantify orders of magnitude, and present possible trade-offs to policymakers, "such as: Given what is happening to your oil reserves, if you do not do something to raise the savings rate...your real incomes will start dropping in five years" [Harberger et al., 1992, 3]. Alternatively, if better methodologies for evaluating CGE models were developed, and particular CGE models stood up well to these evaluations, analysts might be justified in placing more reliance on those models' simulations.

Theoretical Analyses

If empirical models provide an analyst with limited information about the impact of structural changes, can theoretical arguments be used to derive the expected effects? Krueger identifies some important difficulties with this approach:

> A basic problem with micro issues is that theory often predicts what might happen under some conditions, and then someone proposes what should be without thoroughly investigating whether these conditions prevail. This "might/should" linkage destroys a lot of what otherwise might be reasonably careful thinking on policy. [Harberger et al., 1992, 1]

Krueger gives "infant industry" arguments as an example; people argue that the particular instance might involve an infant industry, but provide no criteria for determining whether the industry in question is in fact an "infant." This problem of assessing the applicability of a microeconomic model to a *particular* situation, arises in an interesting way in the case of airline deregulation.

Airline Deregulation. Prior to deregulation economists used theoretical analyses "in ways that suggested that performance without regulatory intervention would approximate perfect competition" [Levine, 1987, 400]. The analysis particularly emphasized that prices would be lower than they were under regulation. These predictions about prices were absolutely correct. However, neither the prederegulation use of the theory of perfect competition nor the post-regulation employment of contestability theory were able to predict the ensuing structure of the airline industry correctly.

The predictions about the post-regulation structure of the industry included forecasts that there could be as many as 200 airlines each operating as few as six aircrafts. Prices would be lower and uniform for all customers, with perhaps some discounts for off-peak flights, and the airlines existing at the time of deregulation would have difficulty competing with the new entrants. Entry and exit would be easy since aircraft were mobile assets.

Levine indicates that significant deviations from the predictions of both the perfect competition and contestability theories have been observed. The structure of the industry is clearly not consistent with the predictions of pure or perfect competition; there are only a small number of large firms, with three dominant.

Some of the other outcomes are also inconsistent with the theoretical expectations. Most of the firms existing in 1978 are still operating. While some of the old firms, such as Braniff, Eastern and Pan Am, exited, most of the new entrants also have failed. Additional deviations from the theoretical predictions include: mergers, vertical integration, hub and spoke systems with hub domination, a complex fare structure with many types of restrictive fares, frequent flyer programs, the use of travel agents and computer reservation systems. These phenomena "cannot adequately be explained by the traditional models of airline competition" [Levine, 408].[12]

Why were the predictions and the outcomes so disconnected? Moreover, why were these outcomes such surprises to those who made the predictions? The answer to the first question is that one of the fundamental theoretical assumptions of contestability theory was not applicable to the airline industry. Specifically, contestability theory implies that the potential "hit and run" tactics of possible entrants will constrain the behavior of incumbents in that market. This situation would produce results similar to those prevailing in a perfectly competitive market. A fundamental assumption of contestability theory is that the response time of incumbents is longer than the length of the period required for new entrants to achieve a costless exit [Spence, 1983]. Contrary to this assumption, the computerized reservations systems of incumbent airlines permit virtually instantaneous pricing responses, while new entrants would have some fixed commitments which would take time to liquidate. With hindsight the aforementioned assumption clearly does not hold in the case of airlines.[13] This case thus serves to illustrate the problems that can be involved in using a theory which had not been previously applied to a specific industry to determine the effects that a major policy change would have on that industry.

However, not all industry observers were "surprised" by the outcomes. The business literature and industry officials, but not the academic literature, indicated that: (1) the airlines viewed their existing route structures in terms of hubs and spokes; (2) the airlines could use their sophisticated computerized reservations sytems to establish non-uniform prices for different types of customers; and (3) that the industry would become more concentrated.[14] While it is dangerous to look for "correct predictions" after an event has occurred, these "forecasts" were, in fact, available at the time that the theoretical studies were undertaken but were not explicitly introduced into the analyses. If such industry views had been considered in the analyses, the number of surprises might have been a smaller. For this reason Stekler [1995] has suggested using analytical business-game simulations, with knowledgeable industry observers as participants, as another procedure for forecasting the effects of structural changes. This would be especially desirable when theoretical and business analyses differed.

CONCLUSIONS

This paper has been concerned with an issue that confronts every policy analyst and forecaster: how to obtain and interpret the information required to implement policies and issue forecasts. The paper indicated that different problems might arise

if data were sought for situations in which the economic structure was changing, as compared to when it was constant. Nevertheless, in both situations the analyst might experience important difficulties in obtaining the relevant and "correct" information. These difficulties encompassed data revisions, conflicting results, publication biases, and possibly inappropriate models and theories. In some cases there are procedures that an analyst can employ to help overcome these difficulties, while in other situations it may be impossible to obtain the needed information. The user/evaluator of these analytical products should be aware of these limitations and take them into account when judging the value of these products.

NOTES

The authors wish to thank Bryan Boulier, two anonymous referees, and the editor of this *Journal* for helpful comments.

1. There is an extensive literature dealing with these data problems. Citations are available from the authors.
2. As one example, a study of inventory investment equations showed that when these structural relationships were estimated with revised data, equation coefficients differed from those obtained for a comparable time period using the early data [Stekler, 1970].
3. For an example of the application of meta-analysis to an economics topic, see Card and Krueger [1995].
4. See, for example, Card and Krueger [1995] and Goldfarb [1995a;1995b] in economics; Berlin, et. al. [1989] and Gelber and Goldhirsch [1991] in biostatistics/medical research; and Lipsey and Wilson [1993] in psychometrics/education research.
5. Complaints of this sort can be found in Leamer [1978; 1983, 37-8], Johnson [1971, 2,10], Denton [1988], and Goldfarb [1995a; 1995b].
6. The biostatistics literature indicates that reliance on published studies can lead to dangerously mistaken inferences: disease treatments that appear efficacious based on *published* clinical trials sometimes turn out to be ineffective when *unpublished* trials are included [Gelber and Goldhirsch, 1991].
7. As an example, see Klein [1991].
8. Reconciling such diverse estimates is analogous to meta-analysis. However, to the best of our knowledge, "meta-analysis-like" techniques have not been developed for reconciling such large-model estimates.
9. Lau indicates that a problem with this "calibration" approach "has to do with the lack of a measure of the degree of reliability of the model and its parameters...The potential consumer of the applied general equilibrium model must be reasonably convinced of the reliability, stationarity, and stability of the model and its parameters before he can have any degree of confidence in its results. With the 'calibration' approach, no measures of the reliability of the model and its parameters are available" [1984, 134-35].
10. "(I)t is quite extraordinary not only how little we know about the numerical values of elasticities... but how (the) little we think we know changes as quickly as it does. In the savings area, for instance, 10 years ago, elasticities were thought to be small, 5 years ago they were thought to be large, and now once again they are thought to be smaller" [Whalley, 1985, 27].
11. A U.S. Congress Joint Economic Committee study [1993] to assess the debate about economic effects of NAFTA examined 16 empirical studies estimating NAFTA's economic effects. Nine of the studies were based on CGE models. The JEC points out that the potential effects of NAFTA depend on the assumptions of the models. The assumptions differed in crucial ways that were likely to affect results about the effects of NAFTA. As one example, seven of the nine CGE models do not allow any change in total U.S. employment; "thus, by definition, these models cannot predict that NAFTA would result in a net reduction in U.S. jobs" [1993, 12].

12. For documentation that a number of industry outcomes are incompatible with contestability theory see Levine [1987] and Evans and Kessides [1993]. For additional discussion of the deregulated industry, see Borenstein [1992].
13. Contestability theory was originally developed for application to telecommunications. The assumption that the incumbent could not immediately respond to a "hit and run" entrant was plausible when applied to the regulatory constraints under which A T and T functioned. However, in the deregulated airline industry, an incumbent carrier with a sophisticated computer system and a complex fare structure could use that computer system to rapidly and selectively change particular fares in response to a specific new entrant on a particular route.
14. For examples of these industry analyses, see *Aviation Week*, [17 February 1975, 20-22; 7 April 1975, 21; 16 February 1976, 7; 29 November 1976, 7; 15 August 1977, 24].

REFERENCES

Becker, B. Combining Significance Levels, in *The Handbook of Research Synthesis* edited by H. Cooper and L. Hedges. New York: Russell Sage, 1994, 215-30.

Berlin, J., Begg, C. and Louis, T. An Assessment of Publication Bias Using a Sample of Published Clinical Trials, *Journal of the American Statistical Association*, June 1989, 381-92.

Borenstein, S. The Evolution of U.S. Airline Competition, *Journal of Economic Perspectives*, Spring 1992, 45-73.

Card, D. and Krueger, A. Time Series Minimum Wage Studies: A Meta-Analysis, *American Economic Review, Papers and Proceedings*, May 1995, 238-43.

Cooper, H. and Hodges, L., eds. *The Handbook of Research Synthesis*. New York: Russell Sage.

Coughlin, C. What Do Economic Models Tell Us About the Effects of the U.S.-Canada Free Trade Agreement? *Federal Reserve Bank of Saint Louis Review*, September/October 1990, 40-58.

Denton, F. The Significance of Significance: Rhetorical Aspects of Statistical Hypothesis Testing in Economics, in *The Consequences of Economic Rhetoric*, edited by A. Klamer, D. McCloskey, and R. Solow. Cambridge: Cambridge University Press, 1988, 163-83.

Evans, W. and Kessides, I. Structure, Conduct and Performance in the Deregulated Airline Industry. *Southern Economic Journal*, January 1993, 450-67.

Gelber, R. and Goldhirsch, A. Meta-Analysis: The Fashion of Summing Up Evidence. *Annals of Oncology*, 1991, 461-68.

Goldfarb, R. If Empirical Work in Economics is Not Severe Testing, What Is It? in *Measurement, Quantification, and Economic Analysis*, edited by I. Rima, London: Routledge, 1995a, 333-64.

_____. The Economist-As-Audience Needs a Methodology of Plausible Inference. *Journal of Economic Methodology*, December 1995b, 21-42.

Harberger, A., Krueger, A. Fox, J. and Edwards, S. Interface Between Economic Techniques and Economic Policy. *Contemporary Policy Issues*, July 1992, 1-15.

Johnson, H. The Keynesian Revolution and the Monetarist Counter-Revolution. *American Economic Review Papers and Proceedings*, May 1971, 1-14.

Klein, L. ed. *Comparative Performance of U.S. Econometric Models*. New York: Oxford University Press, 1991.

Koopmans, T. Economics Among The Sciences. *American Economic Review*, March 1979, 1-13.

Lau, L. Comments, in *Applied General Equilibrium Analysis*, edited by H. Scarf and J. Shoven. Cambridge: Cambridge University Press, 1984, 127-37.

Leamer, E. *Specification Searches: Ad Hoc Inference with NonExperimental Data*. New York: John Wiley and Sons, 1978.

_____. Let's Take the Con Out of Econometrics, *American Economics Review*, March 1983, 31-43.

Levine, M. Airline Competition in Deregulated Markets: Theory, Firm Strategy, and Public Policy. *Yale Journal on Regulation*, 1987, 393-494.

Lipsey, M. and Wilson, D. The Efficacy of Psychological, Educational and Behavioral Treatment-Confirmation from Meta-Analysis. *American Psychologist*, 1993, 1181-209.

Lucas, R. Econometric Policy Evaluation: A Critique, in *The Phillips Curve and Labor Markets, Carnegie-Rochester Conference Series on Public Policy*, edited by K. Brunner and A. Meltzer. *Journal of Monetary Economics*, Supplementary issue, 1976. 19-46.

Spence, M. Reviewing *Contestable Markets and the Theory of Industry Structure. Journal of Economic Literature*, September 1983, 981-90.

Stekler, H. *Economic Forecasting* . New York: Praeger Publishers. 1970.

_____. Forecasting for Large Policy Changes, mimeo, 1995.

U. S. Congress, Joint Economic Committee. Potential Economic Impacts of NAFTA: An Assessment of the Debate, Staff Study, October 1993

Whalley, J. Hidden Challenges in Recent Applied General Equilibrium Exercises, in *New Developments in Applied General Equilibrium Analysis*, edited by J. Piggot and J. Whalley. Cambridge: Cambridge University Press, 1985, 25-41.